Different Types of Data and the Validity of Democracy Measures

Skaaning, Svend-Erik

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Skaaning, S.-E. (2018). Different Types of Data and the Validity of Democracy Measures. *Politics and Governance*, *6*(1), 105-116. https://doi.org/10.17645/pag.v6i1.1183

COGITATIO

Politics and Governance (ISSN: 2183–2463)
2018, Volume 6, Issue 1, Pages 105–116
DOI: 10.17645/pag.v6i1.1183

Article

Different Types of Data and the Validity of Democracy Measures

Svend-Erik Skaaning

Department of Political Science, Aarhus University, 8000 Aarhus, Denmark, E-Mail: skaaning@ps.au.dk

Submitted: 27 September 2017 | Accepted: 20 November 2017 | Published: 19 March 2018

Abstract
Different measures of democracy rely on different types of data. Some exclusively rely on observational data, others rely on judgement-based data in the form of in-house coded indicators or expert surveys. A third set of democracy measures combines information from indicators based on different types of data, some of them also data from representative surveys of the mass public. This article discusses the advantages and disadvantages of these different types of data for the measurement of electoral and liberal democracy. The discussion is based on the premise that the main priorities must be to establish a high degree of concept-measure consistency, i.e. indicators capture relevant aspects of the core concept of interest in a precise and unbiased manner, and to provide high coverage. The basic argument of the article is that no type of data is superior to others in all respects. The article draws on examples from extant datasets to illustrate the tradeoffs and it offers suggestions about how to reduce some of the potential drawbacks.

Keywords
democracy; measuring democracy; reliability; types of data; validity

Issue
This article is part of the issue "Why Choice Matters: Revisiting and Comparing Measures of Democracy", edited by Heiko Giebler (WZB Berlin Social Science Center, Germany), Saskia P. Ruth (German Institute of Global and Area Studies, Germany), and Dag Tanneberg (University of Potsdam, Germany).

> Not everything that can be counted counts,
> and not everything that counts can be counted.
> (Cameron, 1963, p. 13)

1. Introduction

The construction and use of measures of democracy in social scientific research has increased considerably in recent decades. This makes good sense; without them, the identification of trends in political rights and liberties must be based on rough impressions not allowing for systematic temporal and cross-country comparisons (Bollen, 1992, p. 189). However, such efforts are only valuable if the quality of the data is high in terms of reliability and validity.[1]

When we attempt to measure democracy, the identification of empirical indicators that tap into the different aspects of the overarching concept is one of the most important tasks. One can either use extant indicators, collect new data, or combine new indicators with old ones. The main priority must be to establish a high degree of concept-measure consistency, i.e. the extent to which the indicators capture all of the components of the core concept of interest (and only those), and the extent to which they do so in a precise and unbiased manner (Adcock & Collier, 2001; Goertz, 2006; Munck, 2009).[2] In the

[1] *Reliability* concerns whether a measurement procedure produces similar results under consistent conditions. *Validity* concerns the extent to which a measure plausibly captures the concept it is supposed to measure. *Reliability* is a necessary but insufficient condition for measurement validity. See Seawright and Collier (2014) for an overview and critical assessment of different validation strategies applied to measures of democracy. The strategies they discuss mainly apply to extant measures, while they neither discuss the data generating procedures nor address the question of different data types in the same level of detail as the present article.

[2] Note that concept-measure consistency, besides the use of adequate indicators, also concerns the aggregation procedures used to combine the information provided by different indicators. However, the question of whether the aggregation of information provided by the indicators is based on theoretically justified, empirically sound procedures is not part of this article's agenda as it constitutes a rather independent issue (see Bollen & Lennox, 1991; Goertz, 2006; Møller & Skaaning, 2011, Appendix; Munck, 2009).

words of the Office of the High Commissioner for Human Rights (OHCHR, 2012, p. 50):

> An important statistical consideration in identifying and developing human rights indicators, or any set of indicators for that matter, is to ensure their relevance and effectiveness in measuring what they are supposed to measure. This relates to the notion of indicator validity. It refers to the truthfulness of information provided by the estimate or the value of an indicator in capturing the state or condition of an object, event, activity or an outcome for which it is an indicator. Most other statistical and methodological considerations follow from this requirement.

Among the supplementary—and related—criteria that scholars take into consideration are: Whether indicators are produced through transparent and replicable data-generating processes, whether they are made publicly available, and whether they have extensive coverage in terms of units (typically countries) and time (typically years). Researchers face numerous tradeoffs when trying to fulfill these criteria.

One of the most important considerations is what type of data the ever-growing industry of measuring democracy, governance, and human rights should rely on (see Arndt & Oman, 2006; Landman & Carvalho, 2009, Chapter 3; OHCHR, 2012; Schedler, 2012; United Nations Development Programme, 2012).

Different measures of democracy are based on different types of data. Four main data types have been used to construct the major democracy measures: observational data (OD), i.e. data on directly observable facts, such as turnout rates or the presence or absence of formal political institutions; 'in-house' coding (IC) by researchers and/or their assistants based on an assessment of country-specific information found in reports, academic works, newspapers, archival material, etc.; expert surveys (ES), where selected country experts provide an evaluation based on their case-specific knowledge; and representative surveys (RS), where a sample of ordinary citizens provide judgements about particular issues.[3]

All of these types of sources have different strengths and shortcomings. Even though this is well-known, contrasting views about what kind of data is better still exist. To illustrate, the Office of the United Nations High Commissioner for Human Rights (OHCHR, 2012), which represents the global commitment to universal ideals of human dignity, takes a clear stand in favor of observable data in its widely cited report on human rights measurement. This preference for fact-based quantitative indicators over judgement-based indicators[4] is motivated by an interest in making assessments less subjective and thus more broadly acceptable. According to Cheibub, Gandhi and Vreeland (2010, p. 77), the data required by judgement-based democracy measures 'are hard, if not impossible, to obtain. Consequently, we suspect that these measures entail coding created on the basis of inferences, extensions, and perhaps even guesses' (see also Merkel et al., 2016; Przeworski, Alvarez, Cheibub, & Limongi, 2000; Vanhanen, 2000).

In contrast, the people behind the Worldwide Governance Indicators state that fact-based indicators are insufficient for capturing the realities of governance outcomes on the ground (Kaufman & Kraay, 2008). They therefore consider judgement-based data as a valuable tool. This position is motivated by the assumption that it is virtually impossible to capture the relevant aspects of governance, including democracy, without relying on the judgement of experts, in-house coders, and/or citizens (see also Bowman, Lehoucq, & Mahoney, 2005; Coppedge, Gerring, Lindberg, Skaaning, & Teorell, 2017a, 2017b; Munck, 2009; Schedler, 2012).

To increase the awareness among producers and users of democracy data, it seems pertinent to critically review and supplement the arguments and suggestions in a single article. More particularly, this article discusses the pros and cons of different data types and suggests how to counter some of the potential problems related to the measurement of electoral democracy (i.e. access to government power is determined by competitive and inclusive elections) and liberal democracy (i.e. electoral democracy combined with respect for civil liberties and the rule of law) (see Møller & Skaaning, 2011). The discussion draws on extant as well as suggested indicators to illustrate the tradeoffs. After presenting an overview of what kind of data extant democracy measures are based on, I discuss—for each of the four types of data in turn—the potential advantages and disadvantages regarding reliability and validity together with suggestions to reduce some of the problems. The basic argument of the article is that no type of data is superior to the others in *all* respects. Researchers should generally pay more attention to different ways of increasing valid measurement, including the combination of different types of data and data from different sources, whenever they construct their measures. It is not reasonable simply to stick to conformist practices and dogmatic doctrines about the general superiority of one type of data.

2. Extant Democracy Measures: What Kinds of Data Are Used?

Table 1 makes clear that there is considerable variation regarding how many kinds and which kind of data

[3] In this article, I exclusively focus on different types of standards-based data and thus disregard different types of events-based data.

[4] The distinction between fact-based and judgement-based indicators 'refers to information content of the indicators in question. Accordingly, objects, facts or events that can, in principle, be directly observed or verified, such as formal political institutions, are categorized as objective [fact-based] indicators. Indicators based on perceptions, opinions, assessment or judgements expressed by individuals are categorized as subjective [judgement-based] indicators' (OHCHR, 2012, p. 17). However, regarding the measurement of democracy and other governance-related concepts, it is often difficult to make a clear-cut distinction.

sources they build on. This plethora of approaches indicates that it not obvious what kind of data—or mix of data—one should prefer when trying to measure democracy. For some indicators, it is not easy to say if they are fact-based or judgement-based (more on this below). But if we take the statements of the different data providers as given, the Democracy–Dictatorship dataset (Cheibub et al., 2010) and Vanhanen's (2000) polyarchy measure only use observational data. The first of these measures uses indicators of legislative and executive elections, status of the legislature, opposition parties, and government turnovers to create a dichotomous distinction between democracies and autocracies. The second only uses share of votes cast for the largest party and electoral turnout rates in national elections to capture the level of democracy.

The underlying data of the Bertelsmann Transformation Index (Bertelsmann Stiftung, 2017), the Freedom in the World survey (Freedom House, 2017), and the Perception of Electoral Integrity index (Norris, Frank, & Martínez i Coma, 2014) are all based on expert assessments. The Polity Measure (Marshall, Gurr, & Jaggers, 2016) and the CIRI Human Rights Dataset (Cingranelli & Richards, 2010) solely rely on in-house coded data. The remaining measures included in the overview presented in Table 1 build on more than one kind of data source. The Democracy Barometer dataset (Merkel et al., 2016), the Unified Democracy Scores (Pemstein, Meserve, & Melton, 2010), and the Worldwide Governance Indicators (Kaufman & Kray, 2017) do not provide original data collection but use extant indicators based on all four kinds of data sources. The Varieties of Democracy (Coppedge et al., 2017b) dataset relies on all types of data apart from representative surveys, and the Democracy Index by the Economist Intelligence Unit (2007) only excludes in-house coded data. Finally, the Lexical Index of Electoral Democracy (Skaaning, Gerring, & Bartusevičius, 2015) combines two kinds of data sources: in-house coded data and observational data. It varies quite a bit from measure to measure whether the

Table 1. Selected characteristics of 13 large-scale democracy datasets. Source: Coppedge et al. (2017a, p. 6) and own assessment.

Names of Data Provider and Dataset	Years covered	Types of sources				Based on various datasets	Uncertainty estimates
		IC	OD	ES	RS		
Bertelsmann Stiftung (2017): Bertelsmann Transformation Index (BTI)	2003–2015 (biennial)			X		No	No
Cheibub et al. (2010): Democracy–Dictatorship (DD)	1946–2008		X			No	No
Cingranelli & Richards (2010): CIRI Human Rights Database (CIRI)	1981–2011	X				No	No
Coppedge et al. (2017b): Varieties of Democracy dataset (V-Dem)	1900–2016	X	X	X		No	Yes
Economist Intelligence Unit: Democracy Index (EIU)	2006, 2008, 2010–2016		X	X	X	Yes	No
Freedom House: Freedom in the World (FH)	1972–2016			X		No	No
Kaufmann & Kray (2017): Worldwide Governance Indicators (WGI)	1996, 1998, 2000–2015	X	X	X	X	Yes	Yes
Marshall, Gurr, & Jaggers (2016): Polity IV (Polity)	1800–2015	X				No	No
Merkel et al. (2016): Democracy Barometer (DB)	1990–2014	X	X	X	X	Yes	No
Norris et al. (2014): Perceptions of Electoral Integrity (PEI)	2012–2016			X		No	Yes
Pemstein et al. (2010): Unified Democracy Scores (UDS)	1946–2012	X	X	X	X	Yes	Yes
Skaaning et al. (2015): Lexical Index of Electoral Democracy (LIED)	1800–2016	X	X			No	No
Vanhanen (2000): Polyarchy Dataset (Vanhanen)	1810–2014		X			No	Yes

different types of data are used to measure the same subcomponents and components of the overall democracy measures.

Disagreements about best practices regarding what kind of data to employ continue to flourish. The great variation not only reflects differences in resources; it also indicates different weighting of the potential problems related to data types. But what are the more specific pros and cons of different data types? How can the data type choice matter for reliability and validity?

3. Potential Advantages and Disadvantages of Different Kinds of Data

3.1. Observational Data

Observational data have a high, often preferred standing among users of social science data. In the words of Cheibub et al. (2010, p. 74), 'The reliability of a measure depends on whether knowledge of the rules and the relevant facts is sufficient to unambiguously lead different people to produce identical readings on specific cases'. On this basis, they prefer democracy measures based on directly observable and verifiable indicators rather than subjective and fuzzy indicators. Among the main assets of a fact-based approach to measurement are transparency and a replicable data-generation process, which is generally less susceptible to biases than judgement-based data types (see below). Moreover, observational data often provides scales of the phenomena in question that are both relatively easy to interpret and comparable across countries and over time (OHCHR, 2012).

However, the assumptions underlying this preference are criticized for being unrealistic. According to Schedler (2012, p. 28), the collection and use of non-judgmental data in the social sciences rests on two conditions: '(1) transparent empirical phenomena whose observation do not depend on our judgmental faculties and (2) complete public records on those phenomena'. When we want to measure democracy, none of these criteria are met. Not all aspects of democracy are easily observable and, relatedly, official statistics do not capture many relevant features in meaningful ways. Readily observable empirical information is often incomplete, inconsistent, or insufficient. 'Some empirical phenomena we cannot observe in principle, others we cannot observe in practice' (Schedler, 2012, p. 28).

A particular problem emerges when measuring democracy by examining the official (formal) laws of the land, first and foremost the constitution:[5] There is often a large discrepancy between what appears on the books and what is practiced on the ground. Informal rules and traditions are often more important than formal regulations. To illustrate this point with an extreme example, the Soviet 1936 constitution (aka. the Stalin Constitution) promised free and fair elections and respect for civil liberties on top social and economic rights. In practice, however, the political regime was totalitarian (see Linz, 2000), including a level of state repression that has hardly been matched by any other political regime in world history.

This problem refers to more than discrepancies between *de jure* and *de facto* regulations. For example, OHCHR (2012, p. 97) has suggested using reported cases of killing, disappearances, detention, and torture against journalists to measure freedom of opinion. This could be a relevant indicator but has two significant shortcomings. On the one hand, there is likely to be a reporting bias because reliable information is often not readily available (see Fariss, 2014; McNitt, 1988, pp. 94–99; Weidmann, 2016). The perpetrators normally have a clear interest in keeping the correct number secret and it is often difficult to know why a particular journalist has disappeared or died, or whether they were imprisoned due to a legitimate use of their freedom of expression or some other reasons. On the other hand, anticipated sanctions often lead to self-censorship. Journalists are rarely killed in North Korea (as far as we know), because they know that criticizing the government would have dire consequences. These problems would apply to similar attempts at capturing respect for liberal rights and adherence to the rule of law by the (exclusive) use of observational indicators.

Among the attempts to measure democracy using observational data, we find the democracy–dictatorship dataset (Cheibub et al., 2010; Przeworski et al., 2000). Its reliance on the rule of electoral government turnover to determine whether elections have been free suffers from two problems. First, the so-called Botswana problem, i.e. a government seems to be continuously reelected through free elections, meaning that Botswana (and other such cases) does not fulfill the turnover-criterion, saying that an alternation in government power has taken place under electoral rules identical to those bringing the incumbent into power. Second, the turnover criterion is implemented in a way that could introduce further problems. The coding rule says that a government turnover implies that a particular regime is coded as democratic all the way back to when the previous government took power given that the case also fulfilled the other criteria for democracy in the period, if the electoral rules are identical. However, a judgement call is sometimes needed to determine what counts as electoral rules and what counts as a relevant change to these rules (Knutsen & Wig, 2015, p. 909).[6]

The freeness and fairness of elections could also improve significantly (from no uncertainty to significant uncertainty about the outcome) under the same (formal) electoral rules. This applies, among other cases, to the

[5] This is a prominent feature of the Democracy Barometer and a number of governance measures, such as the Rule of Law Index by the World Justice Project (2016).

[6] This point applies more generally to seemingly fact-based indicators used to measure democracy, where elements of judgement cannot be fully excluded.

Dominican Republic between 1966 and 2002. In this period, election outcomes varied greatly and, according to comprehensive case studies, did not meet the minimum threshold for electoral democracy before 1978 (Marsteintredet, 2009, Chapter 4). In other cases, government turnover merely signifies that the ruling coalition is split and no longer controls the sufficient means to stay in power—a situation that the opposition exploits to gain power through manipulated elections. This problem applies to, for example, the change from conservative to liberal hegemonic rule in Columbia (1930–1931) and President Kurmanbek Bakiyev's rise to power in connection with the Tulip Revolution in Kyrgyzstan. Hence, government turnover often provides strong and relevant indication of free electoral competition, but it is not unproblematic and undisputable evidence (see Bogaards, 2007; Boix, Miller, & Rosato, 2013; Skaaning et al., 2015).

Another well-known example is Vanhanen's (2000) use of voter turnout and the vote share of the largest party in order to capture different degrees of democracy. These indicators tend to fail to tap all of the relevant aspects of democracy, however, such as the degree of freedom of expression and the power of the parliament, while capturing things that do not directly reflect the level of democracy, such as mandatory voting, dissatisfaction with the government, and the weather on voting day (Bollen, 1990, pp. 8, 15). Both the official statistics on turnout rates and vote share could also be unreliable—either because the data has been manipulated or because the data providers have been unable to collect all of the relevant information and aggregate them correctly. Some governments simply do not have the capacity to collect and handle the relevant information, which leads to missingness or flawed estimates. Other governments and/or their agents have strong incentives (and few constraints) to manipulate official data in order to misinform their own citizens and foreign governments and organizations (Herrera & Kapur, 2007).

Both of these circumstances can seriously reduce the availability and quality of data that could be relevant for measuring democracy, since they tend to be politically sensitive. Even in the case of so-called hard economic data (e.g. GDP per capita and trade), where governments and international organizations invest extensive resources in the collection of information and calculate the figures used by countless social scientists, there are remarkable problems regarding reliability and validity (Jerven, 2013; Kerner, 2014). There is therefore good reason to refrain from buying into the claim that fact-based data are always more informative and less biased than judgement-based data (see, e.g., Bollen, 1992; Coppedge et al., 2017a, 2017b; Kaufman & Kraay, 2008; Schedler, 2012). Public statistical information and other types of observable data can be useful for measuring democracy, but directly observable indicators do not capture all aspects of democracy well.

3.2. In-House Coded Data

One way of overcoming problems related to the lack of good observable indicators is to base scores on different kinds of relevant information found in diverse sources providing country-specific information, such as newspapers, election observation reports, human rights reports, and academic works. The construction of in-house coded data normally follows a particular procedure: Relevant information is gathered, after which a coder evaluates the evidence on one or more particular issues and translates the evaluation into a score based on more or less explicit and precise standards. Note, furthermore, that in the case of in-house coding, the coders are not experts on all of the (many) countries (and maybe also not the substantive areas) they assign scores to.

In-house coded data has three major advantages: It can be used to capture important traits that are largely undetectable by observational data (Bollen, 1993, p. 1210; Hadenius & Teorell, 2005, pp. 14–15; Mainwaring, Brinks, & Pérez-Liñán, 2001, p. 61; Munck & Verkuilen, 2002, p. 18). In many cases, bits and pieces of evidence can be put together to create a more general understanding of the actual respect for different democratic rights. On this basis, raters can make an informed estimate of the extent of, say, electoral contestation, freedom of expression, and fair trails, which would otherwise be very difficult to capture in a nuanced manner.

Another positive feature of in-house coded data is that the centralized assignment of scores by one or a few selected coders, ceteris paribus, generally makes for a higher degree of consistency when applying coding criteria. The understandings of concepts and scales will simply be more uniform compared to (more 'decentralized') expert surveys and public opinion surveys. In other words, in-house coding facilitates similar applications of standards across countries, especially if the number of coders is low and they are carefully trained and supervised. The use of multiple coders and inter-coder reliability tests are valuable tools to assess whether the assumptions about consensus among coders are met, i.e. there is consistency in the estimate if the data-generating procedure is repeated by the same or different coders (see Gwet, 2014).[7]

The third potential advantage is that in-house coding facilitates standardized and detailed documentation of why particular observations are assigned certain values. Detailed documentation of the motivation behind the particular scores can obviously be very time-consuming, which is probably why it is not provided in connection to any of the democracy measures based on in-house coding.

There are other reasons for hesitating before accepting values derived from in-house coding. The use of in-house coded data (and judgement-based data more generally) is sometimes rejected with reference to its sub-

[7] Such as those made public in connection to BTI, CIRI, and Polity for a single year and, more appropriately, for a random selection of 10% of the country-years in connection to LIED.

jective nature. In contrast to genuine subjective measures, however, such as data on public attitudes, 'they are not supposed to be subjective, but intersubjective: grounded in public facts and public reasons, defensible in the face of critique' (Schedler, 2012, p. 24). Despite this well-taken qualification, coder-specific biases can still influence the scores in different stages of the coding process (Bollen & Paxton, 1998, 2000):

> First, differential use of sources of information, combined with the filtering of information across the world, could lead to specific judge-centered method factors. Second, judges can process the information available to them in such a way as to differentially weight relevant events or to include irrelevant factors. Finally, the methods of constructing a measure might introduce method effects. (Bollen & Paxton, 2000, p. 64)

In-house coders do not have expert knowledge of all of the countries they code. They must therefore rely on secondary sources, which obviously differ with respect to availability and relevance. Systematic distortion of information is likely as it makes its way from the actual practices and events to the sources of information used by the coders. Accessible data can be ordered according to its informative value. The best situation would be for all relevant information to be available, but this is unrealistic. The following ordering of information therefore applies: recorded, accessible, locally reported, and internationally/foreign reported (Bollen, 1992, pp. 198–199). Movement from the former to the latter resembles a filtering process where some information passes through and some does not.

This process is likely to introduce biases. Filters often tend to be selective in non-random fashions, meaning that the information is neither complete nor representative (Foweraker & Krznaric, 2000, p. 766; Milner, Poe, & Leblang, 1999, p. 420). This is due to differences in the openness of countries, how much international attention they receive (influences by size, language, etc.), ideological preferences of the media, specific agendas of scholarly works and reports, and so forth. While most of the providers of original in-house coding (LIED, Polity, V-Dem) use multiple sources (which are generally unspecified), only CIRI makes use of the Country Reports on Human Rights Practices issued by the US State Department.[8] This fact means that the validity to a very high degree depends on the representativeness and impartiality of a single source, which has been accused of being biased—especially in the early releases (see Innes, 1992; Poe, Carey, & Vazquez, 2001; Qian & Yanagizawa, 2009).[9]

In the next step, raters can introduce random and systematic measurement errors by interpreting the sources differently, either because they based their evaluation of different pieces of relevant or irrelevant information, because they weight the same evidence differently, or because they have different understandings of concepts and scales guiding the coding process.[10] According to Raworth (2001, p. 114), 'The identity of the individuals giving the ratings is inevitably open to questioning'.

Differences in the specific coding processes can also influence the scores. Raters can assign scores to many or few countries (and different groups of countries); they can finalize scores immediately or go back and revise some of them; they can code everything between one year and hundreds of consecutive years at the time; and they can work on the coding in a relatively short but intensive period or carry out the task over a longer, less-intensive period. All of these factors will tend to influence the implicit reference points in the minds of coders and thus have an impact on the scores. The ability of in-house coded data to capture latent regime features in a consistent way is promising, while biases introduced in the coding process and the lack of comprehensive case knowledge are among the potential downsides of this kind of data.

3.3. Expert Survey Data

Expert survey data is generated through assessments of the fulfilment of democratic rights with the help of informed experts, often scholars or other persons working in related fields and intimately acquainted with the subject matter, such as journalists or leading members of NGOs. The main advantage of expert surveys compared to in-house coded data is exactly the case knowledge. The experts presumably know the relevant context and details about the issues in question (Marquardt et al., 2017). If their knowledge is insufficient, they have a superior background for finding relevant information. Experts may even have sufficient contextual knowledge to provide a plausible estimate if there is limited available evidence in terms of written sources directly tapping into a particular phenomenon. Original expert surveys are part of BTI, EIU, FH, PEI, and V-Dem; the three former only use one expert per country, while PEI and V-Dem use multiple experts per country (Coppedge et al., 2017a, p. 8). V-Dem even divides its survey into different categories, and to some degree enlists different experts to fill out different parts of the overall survey for each country (Coppedge et al., 2017b).

The potential problems identified in relation to in-house coded data also apply to expert surveys. The filtering of information might not be as big a problem due to

[8] To code physical integrity rights, CIRI also employs the Annual Reports from Amnesty International.

[9] For detailed discussions of potential biases in the Freedom House scores, see Bollen and Paxton (2000), Giannone (2010), and Steiner (2016).

[10] As stated by Bollen (1990, p. 18), 'A variety of personal factors could unconsciously affect a judge's ratings. These include the relation of the country being rated to the judge's home country, the political orientation of the judge, or any personal stakes in the rating'. Actually, one kind of personal stake, namely academic credibility, will tend to increase the quality of the data, while disinterest in the quality of the product (as is probably the case, at least in relative terms, for many research assistants) can produce low reliability.

the case expertise. However, the selection and weighting of evidence and the coding process will differ somewhat from expert to expert, partly depending on personal factors, such as updated and relevant familiarity with the cases, political leaning, job situation, and work effort (Bollen & Paxton, 1998). Expert knowledge varies and is sometimes inadequate, and the experts often lack strong incentives to enlist and spend much time doing a serious coding job, including searches for additional information. Furthermore, limited and differentiated knowledge leaves room for the so-called 'halo effect,' which is the tendency for a good (or bad) impression of performance in one area to influence opinion regarding other areas (Sequeira, 2012). These circumstances draw attention to the three-fold challenge related to the recruitment of experts. The experts should preferably be the most knowledgeable, unbiased, and be ready to do a careful job. However, the enrolled experts are rarely the best possible according to these criteria.

Experts are also more prone to apply different coding criteria than in-house coders because expert surveys are mostly carried out as decentralized coding without prior training, meaning that the basic understanding of concepts and scales can vary greatly (see Martinez i Coma & Van Ham, 2015; Steenbergen & Marks, 2007). BTI, EUI, and FH combine their expert assessments with review and deliberation across a team of in-house analysts. For good reason, this approach is assumed to increase cross-country consistency. The procedures are not transparent, however, since it is not made public which changes are introduced to the original expert-based values and why for any of the cases.

V-Dem has a different approach to increase the comparability and reduce the influence of potential biases. A complex Bayesian IRT measurement model uses information about agreement across coders, self-assigned uncertainty estimates by the experts about their own ratings, personal coder characteristics (extracted from a post-questionnaire survey), links between countries based on experts assessing more than one country (either for all years or one year), and vignettes related to the survey questions in order to align the experts' thresholds (see Pemstein et al., 2017). This procedure also supplements the scores with a systematic assessment of measurement uncertainty. This is also done for PEI but only based on the degree of expert agreement.

The documentation of the justifications for the scores is desirable, just as in the case of in-house coding. Even though it is usually impossible for experts 'to relate the numerical conclusions they reach to the precise pieces and bits of information that have gone into them…they should be able to document the big picture [and] describe the range of uncertainty and controversy regarding their judgmental decisions with reference to concrete documentary evidence (or the lack of such evidence)' Schedler (2012, p. 32). The extra workload for the experts and coordinators to provide and standardize the information makes this procedure very resource-demanding. Nonetheless, BTI and FH complement their scores with relatively detailed country reports, meaning that one can get an impression of what events and circumstances have influenced the scores for different aspects of democracy (but they do not provide adequate references to the material on which the reports are based).

In sum, the comparative advantage of expert surveys comes to the fore in situations of incomplete or inconsistent information, where contextual knowledge can be used to bridge informational gaps (Schedler, 2012, p. 28). However, the reliance on the personal judgements of a few experts means that the data might lack comparability and might be affected by different kinds of biases.

3.4. Representative Survey Data

The final type of data, representative surveys of the general population, brings the knowledge and opinions of ordinary citizens into play. Mayne and Geissel (2016) argue in favor of including a citizen component in the measurement of democratic quality. It should capture the citizens' democratic commitments, political capacities, and political participation. This perspective, however, seems more relevant for the measurement of deliberative and participatory democracy than electoral and liberal democracy. In connection to these more limited understandings of democracy, the suggested additions are better understood as possible causes or consequences of democracy. Pickel, Breustedt and Smolka (2016) also advocate for the inclusion of representative survey data in the measurement of democratic quality. They propose that citizen evaluations of democratic performance should complement other types of data.

For some purposes, representative surveys can provide valuable information. Respondents can function as 'everyday experts' on issues that are otherwise hard to get firm knowledge about. A case in point is petty corruption, where the experiences of citizens with having to pay bribes could be a superior source of information (see Naval, Walter, & de Miguel, 2008; Razafindrakoto & Roubaud, 2010). Another would be information about whether citizens experience or participate in political violence (see Bhavnani & Backer, 2007).

However, there are also noteworthy problems associated with the use of data from representative surveys to measure democracy. Most citizens lack nuanced knowledge about the general dynamics and performance of particular political institutions. Gut feelings and personal opinions are thus likely to influence the scores. Most drawbacks of judgement-based data apply more strongly to representative survey data than in-house coded data and data based on expert surveys (cf. Marquardt et al., 2017). Experts and in-house coders generally have better backgrounds for carrying out such assessments. They generally possess a broader knowledge regarding the political history of other countries and data collection procedures, a higher degree of shared understanding about

the meaning of particular concepts, and a strong scientific ethos (or least an interest in maintaining their academic credibility). This implies that individual biases and dissimilar standards (both within and across countries) in the interpretation of questions and scales are more pronounced. Ordinary citizens also tend to be more susceptible to collective cultural biases (nation-wide inclinations), and the respondents in representative surveys are very unlikely to provide any form of systematic reasoning for their entries. Ordinary citizens might also be afraid to share their experiences or express their honest opinion, especially in the case of an oppressive regime (Tannenberg, 2017).

Does this mean that we should generally refrain from using representative surveys to measure at all? Ordinary citizens might possibly possess valuable knowledge based on their real-life experiences that could supplement that of experts. Here, it seems pertinent to distinguish between experience-based questions and perception-based questions. The former ask citizens about their own experiences regarding particular situations (e.g. how often they have been asked to pay a bribe or been subjected to violent assaults in the previous year). The latter is typically based on more abstract questions, asking about the lay of the land regarding democracy, civil liberties, corruption, etc.

The experience-based questions have greater potential for providing relevant information than the second type, which are likely to produce unreliable and biased democracy indicators. Combined with the relatively low coverage in terms of years and countries,[11] it is therefore unadvisable to use perception-based data from representative surveys for democracy measurement. None of the evaluated measures are based on original data collection using this approach, but DB, EIU, UDS, and WGI rely on such data—either directly or indirectly (by including composite measures that use them).

4. Discussion

There are several ways of countering the disadvantages identified above. In relation to in-house coded and expert survey data, the documentation should ideally provide answers to the following questions: What evidence has been used and why? And how has the evidence been weighed and processed and why? That is, the criteria for identifying and selecting relevant sources and the criteria for extracting and using relevant information must be pinned down. This work can be done to different degrees of perfection to the point where every score is supplemented with nuanced description of the evidence (using active citation; see Moravcsik, 2014), how and why it has been weighted in certain ways (with relevance as the main criteria; see Bowman et al., 2005; Lustik, 1996; Møller & Skaaning, 2017), and who has been involved and how in the data collection and processing (Schedler, 2012, p. 33).

Inconsistency and personal biases can be reduced by the construction and application of specific and justified definitions of what one attempts to measure and the scales used to distinguish between different levels of fulfilment. The clarification should preferably be presented as precisely as possible and linked to concrete (maybe even paradigmatic) examples. This would support the establishment of shared anchors for the assignment of values. Another useful tool is to reduce conceptual complexity through disaggregation. This would imply the coding of more concrete issues than just freedom of expression, including media censorship, freedom of private discussion, harassment of journalists, and monopoly of news media.

Other factors, such as the exposure of coders to extensive relevant variation, can also improve the consistency. As a rule of thumb, they are more likely to employ similar standards across and within cases when the following conditions are fulfilled: The coders assign scores to a diverse set of many countries; they are willing and allowed to revise scores; they score long time series; and they score the cases within a relatively short period.

If in-house coders or experts score the same cases, formal measurement models can produce replicable point estimates and estimates of uncertainty.[12] One should note, however, that whereas it will almost always be good to increase the number of in-house coders (although there will be a diminishing return), more is not always better in the recruitment of expert coders because there will be a rather limited number of people with high levels of relevant expertise. Moreover, an increase in the number of coders will increase the costs attached to the data collection, thereby emphasizing the latent tradeoff between high quality data and coverage.[13]

Formal measurement models can also be used to combine data from different datasets based on different data-generating approaches (i.e. observational data, in-house coded data, expert survey data, and/or representative survey data) (Bollen & Paxton, 2000, p. 79). The advantage of such composite measures is their utilization of information from several variables. The combination of information from different data types can increase the ability to capture related, but distinct, aspects of the variable in question. In addition, it can reduce the impact of idiosyncratic measurement errors associated

[11] In most cases, it is overly demanding to request respondents to answer questions for several years, coding back in time. Moreover, for different reasons (e.g. regime type, geography, level of socio-economic development), it is extremely difficult to carry out high quality representative surveys in some countries.

[12] Besides the original scores, formal measurement models can utilize other types of information, such as data on the personal characteristics of the experts or in-house coders and their responses to vignettes linked to the variables (see Pemstein et al., 2017). It is also possible to use a measurement model approach to calculate point estimates and uncertainty in the case of representative surveys.

[13] This caveat about a higher demand for resources applies to several of the suggestions, including circumstances where data providers do not themselves possess the relevant skills for implementing them.

with individual indicators. The use of multiple indicators for the same phenomenon also facilitates an assessment of how precise the point estimates are through the construction of confidence levels (see Fariss, 2014; Pemstein et al., 2010). This integrative approach is used (in full or in part) to construct several of the democracy measures (see Table 1). By reducing some problems, however, it risks introducing or increasing others. The integration can lead to an accumulation of the problems associated with the individual indicators rather than resolving them. Moreover, the products tend to be more complex. This means that the relationship between measures and the concepts they should capture becomes more blurred.

Extant democracy measures build on different kinds of data; some only employ in-house coded data, expert survey data, or observable indicators, while others use different combinations of two or more of these types and representative survey data. The identification of the pros and cons typically associated with the respective data types has demonstrated that the different methodological choices about this issue matter for the reliability and validity of democracy measures. Table 2 summarizes some of the most important strengths and weaknesses typically associated with the different data types. Some of the similarities and dissimilarities follow the overall fact-based and judgement-based distinction, while others do not. The overview reveals that the pros and cons associated with the respective kinds of data are not simply mirror images of each other.

The discussion reveals the simplicity of the bullet points in Table 2. They neglect the many nuances of coding rules and processes that can influence the quality of the data. The comparative advantages and disadvantages of the different data types vary in both kind and degree. The reliability and validity depend on the particular procedures used in the data generating process and the aspects of democracy one attempts to capture.

The discussion has also revealed that no type of data is superior to all of the others in all respects when it comes to measuring the fulfilment of democratic rights. Hence, the arguments presented have challenged what many consider conventional wisdom, namely the general superiority on one kind of data—directly observable (fact-based) data. Actually, this belief tends to be a dogmatic doctrine resting on invalid assumptions. As neatly summarized by Schedler (2012, p. 21), 'Banning judgment from measurement is neither a feasible methodological imperative nor a desirable one', and:

> If we were to renounce our judgmental faculties in the measurement of regime properties and regime dynamics, we would have to renounce the measurement of most of the most interesting regime properties and regime dynamics. If we truly had expelled judgment from data development, quantitative research on political regimes could not have blossomed as it has over the past decades. (Schedler, 2012, p. 33)

This point applies more to the measurement of thicker understandings of democracy, such as liberal democracy. Respect for civil liberties and adherence to the rule of law tend to be even harder to capture without judgement-based indicators than narrow electoral criteria (regular, inclusive, and competitive elections). Considerable effort has already been invested in improving democracy measures. More can still be done to increase the reliability and validity, however, and greater awareness about these issues among data users is required.

Acknowledgments

I would like to thank the editors, the reviewers, Carl Henrik Knutsen, Jørgen Møller, and participants in the NOPSA workshop on the Progress and Decline in Democracy Outside the West for valuable comments.

Conflict of Interests

The author declares no conflict of interests.

Table 2. General advantages and disadvantages associated with different data types.

	Advantages	Disadvantages
Observational data	• Avoid personal biases • Fixed and comparable scales • Transparent documentation of scores	• Relevant information often not directly observable • Biases and limitations in available information
In-house coded data	• Consistency in the application of coding criteria • Capture latent traits	• Personal biases • Biases and limitations in the available information • Limited, case-specific knowledge
Expert survey data	• Case-specific knowledge • Capture latent traits	• Personal biases • Inconsistently applied coding criteria
Representative survey data	• Experience-based knowledge • Capture latent traits	• Personal biases • Unfeasible in particular settings • Inconsistently applied coding criteria

References

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, *95*(3), 529–546.

Arndt, C., & Oman, C. (2006). *Uses and abuses of governance indicators*. Paris: OECD Development Centre.

Bertelsmann Stiftung. (2017). *Bertelsmann transformation index*. Retrieved from http://www.bti-project.org/en/home

Bhavnani, R., & Backer, D. (2007). *Social capital and political violence in Sub-Saharan Africa* (Afrobarometer Working Paper 90). Michigan, MI: Michigan State University.

Bogaards, M. (2007). Measuring democracy through election outcomes: A critique with African data. *Comparative Political Studies*, *40*(10), 1211–1237.

Boix, C., Miller, M., & Rosato, S. (2013). A complete dataset of political regimes, 1800–2007. *Comparative Political Studies*, *46*(12), 1523–1554.

Bollen, K. (1990). Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development*, *25*(1), 7–24.

Bollen, K. (1992). Political rights and political liberties. In T. Jabine & R. Claude (Eds.), *Human rights and statistics: Getting the record straight* (pp. 188–213). Philadelphia, PA: University of Pennsylvania Press.

Bollen, K. (1993). Liberal democracy: Validity and method factors in cross-national measures. *American Journal of Political Science*, *37*(4), 1207–1230.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *100*(2), 305–314.

Bollen, K., & Paxton, P. (1998). Detection and determinants of bias in subjective measures. *American Sociological Review*, *63*(3), 465–478.

Bollen, K., & Paxton, P. (2000). Subjective measures of political democracy. *Comparative Political Studies*, *33*(1), 58–86.

Bowman, K., Lehoucq, F., & Mahoney, J. (2005). Measuring political democracy: Case expertise, data adequacy, and Central America. *Comparative Political Studies*, *38*(8), 939–970.

Cameron, W. (1963). *Informal sociology: A casual introduction to sociological thinking*. New York, NY: Random House.

Cheibub, J. A., Gandhi, J., & Vreeland, J. R. (2010). Democracy and dictatorship revisited. *Public Choice*, *143*(1/2), 67–101.

Cingranelli, D., & Richards, D. (2010). The Cingranelli and Richards (CIRI) human rights data project. *Human Rights Quarterly*, *32*(2), 401–424.

Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E., & Teorell, J. (2017a). *V-dem comparisons and contrasts with other measurement projects* (The Varieties of Democracy Institute Working Paper Series 2017: 45). Gothenburg: University of Gothenburg.

Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Krusell, J., . . . Wilson, S. (2017b). *V-dem methodology v7.1*. Gothenburg: Varieties of Democracy (V-Dem) Project.

Economist Intelligence Unit. (2007). *The Economist intelligence unit's index of democracy*. Retrieved from https://www.economist.com/media/pdf/DEMOCRACY_INDEX_2007_v3.pdf

Fariss, C. (2014). Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review*, *108*(2), 297–318.

Foweraker, J., & Krznaric, R. (2000). Measuring liberal democratic performance: An empirical and conceptual critique. *Political Studies*, *48*(4), 759–787.

Freedom House. (2017). About freedom in the world: An annual study of political rights and civil liberties. *Freedom House*. Retrieved from https://freedomhouse.org/report-types/freedom-world

Giannone, D. (2010). Political and ideological aspects in the measurement of democracy: The Freedom House case. *Democratization*, *17*(1), 68–97.

Goertz, G. (2006). *Social science concepts: A user's guide*. Princeton, NJ: Princeton University Press.

Gwet, K. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters*. Gaithersburg: Advanced Analytics.

Hadenius, A., & Teorell, J. (2005). *Assessing alternative indices of democracy* (Committee on Concepts and Methods Working Paper Series, WP 6). Mexico City: The Committee on Concepts and Methods.

Herrera, Y., & Kapur, D. (2007). Improving data quality: Actors, incentives, and capabilities. *Political Analysis*, *15*(4), 365–86.

Innes, J. (1992). Human rights reporting as a policy tool: An examination of the State Department country reports. In T. Jabine & R. Claude (Eds.), *Human rights and statistics: Getting the record straight* (pp. 235–257). Philadelphia, PA: University of Pennsylvania Press.

Jerven, M. (2013). *Poor numbers: How we are misled by African development statistics and what to do about it*. Ithaca, NY: Cornell University Press.

Kaufman, D., & Kray, A. (2017). *Worldwide governance indicators*. Retrieved from http://info.worldbank.org/governance/wgi/index.aspx#home

Kaufmann, D., & Kraay, A. (2008). Governance indicators: Where are we and where should we be going? *World Bank Research Observer*, *23*(1), 31–36.

Kerner, A. (2014). What we talk about when we talk about foreign direct investment. *International Studies Quarterly*, *58*(4), 805–881.

Knutsen, C. H., & Wig, T. (2015). Government turnover and the effects of regime type: How requiring alternation in power biases against the estimated economic benefits of democracy. *Comparative Political Studies*, *48*(7), 882–914.

Landman, T., & Carvalho, E. (2009). *Measuring human rights*. London: Routledge.

Linz, J. (2000). *Totalitarian and authoritarian regimes*. Boulder, CO: Lynne Rienner Publishers.

Lustik, I. (1996). History, historiography, and political science: Multiple historical records and the problem of selection bias. *American Political Science Review*, *90*(3), 605–618.

Mainwaring, S., Brinks, D., & Pérez-Liñán, A. (2001). Classifying political regimes in Latin America, 1945–1999. *Studies in Comparative International Development*, *36*(1), 37–65.

Marquardt, K., Pemstein, D., Petrarca, C. S., Seim, B., Wilson, S. L., Bernhard, M., . . . Lindberg, S. I. (2017). *Experts, coders, and crowds: An analysis of substitutability* (The Varieties of Democracy Institute Working Paper Series 2017: 53). Gothenburg: University of Gothenburg.

Marshall, M. G., Gurr, T., & Jaggers, K. (2016). *Polity IV Project: Political regime characteristics and transitions, 1800–2016*. Vienna, VA: Center for Systemic Peace. Retrieved from http://www.systemicpeace.org/inscr/p4manualv2016.pdf

Marsteintredet, L. (2009). *Political institutions and democracy in the Dominican Republic: A comparative case-study*. Saarbrücken: VDM Verlag.

Martinez i Coma, F., & Van Ham, C. (2015). Can experts judge elections? Testing the validity of expert judgments for measuring election integrity. *European Journal of Political Research*, *54*(2), 305–325.

Mayne, Q., & Geissel, B (2016). Putting the demos back into the concept of democratic quality. *International Political Science Review*, *37*(5), 634–644.

McNitt, A. (1988). Some thoughts on the systematic measurement of the abuse of human rights. In D. Cingranelli (Ed.), *Human rights: Theory and measurement* (pp. 89–103). Basingstoke: Macmillan.

Merkel, W., & Bochsler, D. Bousbah, K., Bühlmann, M., Giebler, H., Hänni, M., . . . Wessels, B. (2016). *Democracy barometer: Methodology. Version 5*. Retrieved from http://www.democracybarometer.org/Data/Methodological_Explanatory_1990-2014.pdf

Milner, W. T., Poe, S. C., & Leblang, D. (1999). Security rights, subsistence rights, and liberties: A theoretical survey of the empirical landscape. *Human Rights Quarterly*, *21*(2), 403–443.

Møller, J., & Skaaning, S.-E. (2011). *Requisites of democracy*. London: Routledge.

Møller, J., & Skaaning, S.-E. (2017). *Reducing bias when enlisting the work of historians: A criterial framework*. Manuscript in preparation.

Moravcsik, A. (2014). Transparency: The revolution in qualitative research. *PS: Political Science & Politics*, *47*(1), 48–53.

Munck, G. (2009). *Measuring democracy: A bridge between scholarship and politics*. Baltimore, MD: John Hopkins University Press.

Munck, G., & Verkuilen, J. (2002). Conceptualizing and measuring democracy: Alternative indices. *Comparative Political Studies*, *35*(1), 5–34.

Naval, C., Walter, S., & de Miguel, R. S. (Eds.). (2008). Measuring human rights and democratic governance: Experiences and lessons from metagora [Special issue]. *OECD Journal on Development*, *9*(2).

Norris, P., Frank, R., & Martínez i Coma, F. (2014). Measuring electoral integrity around the world: A new dataset. *PS: Political Science & Politics*, *47*(4), 1–10.

Office of the High Commissioner for Human Rights. (2012). *Human rights indicators: A guide to measurement and implementation*. Geneva: OHCHR.

Pemstein, D., Meserve, S., & Melton, J. (2010). Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis*, *18*(4), 426–449.

Pemstein, D., Marquardt, K. L., Tzelgov, E., Wang, Y., Krusell, J., & Miriet, F. (2017). *The varieties of democracy measurement model: Latent variable analysis for cross-national and cross-temporal expert-coded data* (The Varieties of Democracy Institute Working Paper Series 2017: 21). Gothenburg: University of Gothenburg.

Pickel, S., Breustedt, W., & Smolka, T. (2016). Measuring the quality of democracy: Why include the citizens' perspective? *International Political Science Review*, *37*(5), 645–655.

Poe, S., Carey, S., & Vazquez, T. (2001). How are these pictures different? A quantitative comparison of the US State Department and Amnesty International human rights reports, 1976–1995. *Human Rights Quarterly*, *23*(3), 650–677.

Przeworski, A., Alvarez, M. E., Cheibub, J. A., & Limongi, F. (2000). *Democracy and development*. New York, NY: Cambridge University Press.

Qian, N., & Yanagizawa, D. (2009). The strategic determinants of US human rights reporting: Evidence from the Cold War. *Journal of the European Economic Association*, *7*(2/3), 446–457.

Raworth, K. (2001). Measuring human rights. *Ethics and International Affairs*, *15*(1), 111–131.

Razafindrakoto, M., & Roubaud, F. (2010). Are international databases on corruption reliable? A comparison of expert opinion surveys and household surveys in sub-Saharan Africa. *World Development*, *38*(8), 1057–1069.

Schedler, A. (2012). Judgment and measurement in political science. *Perspectives on Politics*, *10*(1), 21–36.

Seawright, J., & Collier, D. (2014). Rival strategies of validation: Tools for evaluating measures of democracy. *Comparative Political Studies*, *47*(1), 111–138.

Sequeira, S. (2012). Advances in measuring corruption in the field. *Research in Experimental Economics*, *15*(1), 145–175.

Skaaning, S.-E., Gerring, J., & Bartusevičius, H. (2015). A lexical index of electoral democracy. *Comparative Political Studies*, *48*(12), 1491–1525.

Steenbergen, M., & Marks, G. (2007). Evaluating expert

judgments. *European Journal of Political Research*, *46*(3), 347–366.

Steiner, N. (2016). Comparing Freedom House democracy scores to alternative indices and testing for political bias: Are US allies rated as more democratic by Freedom House? *Journal of Comparative Policy Analysis: Research and Practice*, *18*(4), 329–349.

Tannenberg, M. (2017). *The autocratic trust bias: Politically sensitive survey items and self-censorship* (The Varieties of Democracy Institute Working Paper Series 2017: 49). Gothenburg: University of Gothenburg.

United Nations Development Programme. (2012). *Governance indicators: A user's guide*. Oslo: UNDP.

Vanhanen, T. (2000). A new dataset for measuring democracy, 1810–1998. *Journal of Peace Research*, *37*(2), 251–265.

Weidmann, N. (2016). A closer look at reporting bias in conflict event data. *American Journal of Political Science*, *60*(1), 206–218.

World Justice Project. (2016). *WJP rule of law index 2016*. Retrieved from https://worldjusticeproject.org/our-work/wjp-rule-law-index/wjp-rule-law-index-2016

About the Author

Svend-Erik Skaaning is professor of political science at Aarhus University and co-principal investigator of Varieties of Democracy (V-Dem). His research interests include comparative methodology and the conceptualization, measurement, and explanation of democracy and human rights. He has published numerous articles and books on these issues, including *Requisites of Democracy, Democracy and Democratization in Comparative Perspective*, and *The Rule of Law*.